Heatmap grids comparing DQN, PG, and PPO across passive_env and active_env conditions for Acrobot (Reward = -100) and CartPole (Reward = 500).

DQN — passive_env (Acrobot)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	75.00	58.33	50.00	100.00	83.33	41.67
hd_ly	66.67	50.00	68.75	100.00	83.33	50.00
nr_nb	60.00	50.00	50.00	100.00	100.00	50.00
opti	75.00	50.00	50.00	100.00	75.00	33.33
ln_rt	100.00	100.00	33.33	100.00	75.00	66.67
gamm	66.67	43.75	75.00	100.00	66.67	33.33

DQN — active_env (Acrobot)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	75.00	58.33	60.00	100.00	83.33	50.00
hd_ly	75.00	50.00	56.25	100.00	100.00	56.25
nr_nb	70.00	50.00	33.33	100.00	100.00	56.25
opti	75.00	50.00	50.00	100.00	100.00	33.33
ln_rt	100.00	100.00	50.00	100.00	75.00	33.33
gamm	66.67	62.50	62.50	100.00	83.33	33.33

DQN — passive_env (CartPole)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	50.00	50.00	64.29	60.00	43.75	50.00
hd_ly	64.29	50.00	71.43	66.67	37.50	55.56
nr_nb	64.29	42.86	66.67	70.00	50.00	50.00
opti	80.00	41.67	70.00	50.00	40.00	41.67
ln_rt	43.75	43.75	50.00	80.00	33.33	50.00
gamm	55.56	33.33	55.56	91.67	44.44	50.00

DQN — active_env (CartPole)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	66.67	57.14	57.14	90.00	62.50	44.44
hd_ly	57.14	66.67	71.43	100.00	37.50	55.56
nr_nb	71.43	57.14	66.67	90.00	78.57	66.67
opti	70.00	41.67	70.00	100.00	40.00	58.33
ln_rt	50.00	50.00	50.00	90.00	50.00	50.00
gamm	55.56	55.56	66.67	91.67	66.67	66.67

PG — passive_env (Acrobot)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	66.67	55.56	66.67	71.43	64.29	50.00
hd_ly	61.11	50.00	77.78	64.29	64.29	44.44
nr_nb	66.67	61.11	66.67	68.75	85.71	44.44
opti	64.29	42.86	50.00	66.67	75.00	50.00
ln_rt	85.71	78.57	50.00	58.33	66.67	38.89
gamm	55.56	50.00	61.11	50.00	66.67	50.00

PG — active_env (Acrobot)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	50.00	61.11	66.67	57.14	71.43	61.11
hd_ly	66.67	66.67	72.22	71.43	78.57	50.00
nr_nb	61.11	61.11	66.67	75.00	92.86	50.00
opti	92.86	42.86	50.00	66.67	75.00	50.00
ln_rt	78.57	71.43	50.00	75.00	66.67	61.11
gamm	50.00	50.00	77.78	55.56	55.56	66.67

PG — passive_env (CartPole)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	50.00	38.89	50.00	75.00	66.67	50.00
hd_ly	61.11	50.00	37.50	91.67	64.29	75.00
nr_nb	50.00	43.75	66.67	66.67	42.86	87.50
opti	50.00	41.67	41.67	50.00	50.00	100.00
ln_rt	58.33	42.86	50.00	80.00	66.67	100.00
gamm	50.00	50.00	62.50	83.33	50.00	100.00

PG — active_env (CartPole)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	83.33	66.67	62.50	83.33	83.33	75.00
hd_ly	83.33	50.00	50.00	91.67	85.71	87.50
nr_nb	81.25	68.75	66.67	83.33	78.57	100.00
opti	58.33	66.67	66.67	100.00	70.00	100.00
ln_rt	58.33	71.43	78.57	80.00	66.67	100.00
gamm	87.50	75.00	75.00	100.00	66.67	100.00

PPO — passive_env (Acrobot)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	66.67	50.00	50.00	83.33	61.11	55.56
hd_ly	66.67	66.67	50.00	88.89	61.11	88.89
nr_nb	55.56	50.00	50.00	94.44	61.11	50.00
opti	44.44	55.56	44.44	100.00	61.11	50.00
ln_rt	55.56	66.67	55.56	88.89	66.67	55.56
gamm	50.00	38.89	61.11	72.22	66.67	66.67

PPO — active_env (Acrobot)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	66.67	61.11	61.11	83.33	83.33	77.78
hd_ly	66.67	66.67	61.11	77.78	61.11	83.33
nr_nb	88.89	61.11	66.67	100.00	77.78	77.78
opti	61.11	66.67	55.56	100.00	66.67	50.00
ln_rt	55.56	77.78	66.67	94.44	100.00	66.67
gamm	94.44	61.11	88.89	88.89	83.33	66.67

PPO — passive_env (CartPole)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	33.33	66.67	44.44	50.00	57.14	75.00
hd_ly	50.00	50.00	50.00	66.67	57.14	78.57
nr_nb	50.00	50.00	50.00	50.00	71.43	66.67
opti	50.00	50.00	33.33	75.00	40.00	62.50
ln_rt	57.14	57.14	57.14	70.00	50.00	60.00
gamm	58.33	64.29	50.00	87.50	60.00	50.00

PPO — active_env (CartPole)

	ac_fc	hd_ly	nr_nb	opti	ln_rt	gamm
ac_fc	100.00	72.22	66.67	100.00	78.57	100.00
hd_ly	72.22	66.67	72.22	91.67	71.43	71.43
nr_nb	83.33	72.22	66.67	91.67	71.43	66.67
opti	100.00	66.67	75.00	100.00	90.00	87.50
ln_rt	78.57	78.57	71.43	100.00	100.00	80.00
gamm	91.67	71.43	75.00	100.00	80.00	100.00

Acrobot, Reward = -100   CartPole, Reward = 500